D0 Remote Reprocessing.

D0 has successfully reprocessed over 500 million events using a distributed system that included grid tools and tested grid functionality

News Update – 26th April 2004

Updated: May 6th 2005

Particle Physics Data Grid

The DØ experiment is routinely using the Data Grid to reprocess its recorded data to create datasets for physics analysis. D0 has developed the SAMGrid software, relying on Condor and Globus middleware, to distribute and manage the data.

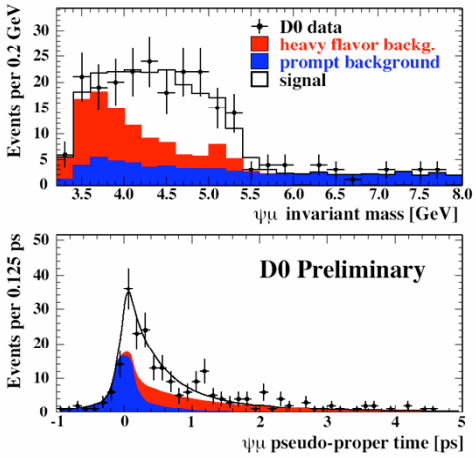
In 2004 D0 used six computing centres in six countries. SAM-Grid has now been deployed to about a dozen sites in US, EU, and Canada. The infrastructure is now being used for the next DZero data reprocessing effort. This consists of 250 TB of physics data that need to be re-processed to take advantage of a better understanding of the detector. To complete the task in time, the effort requires 1600 GHz-CPU year of computing power. Special care has been given to the scalability of the infrastructure at each site: the grid can support running on the order of 1000 concurrent jobs at every site, if the resources are available.

The data is transmitted using Grid file transfer protocols, and Grid job management tools are used to schedule and track the jobs. PPDG collaborative work between D0 and the Condor and Globus teams has provided extensions to and hardening of the software to enable system to perform to the researchers needs.

Benefits to the Science

"The reprocessing effort pushes the limits of our software and infrastructure so that we can get the most physics out of the data collected by the DZero detector," said Dugan O'Neil a participant in the WestGrid collaboration and D0 experiment. "The grid allows DZero to make better use of remote human resources as well as computing power."

DØ has been using and further developing SamGrid for simulation processing and analysis of data acquired from the Tevatron. Fermilab physicist Gustaaf Brooijmans says "Using grids is now an essential and standard part to getting physics out of DØ data from the Tevatron. We could no longer get physics results out without using grid middleware to distribute data and schedule jobs. There is of course a lot left to be done to have grids provide the robustness and universality of our ESnet and other networks - but the potential is proven". The results below show the B(c) mesons mass and lifetime measurement at DØ.



Hardening the Middleware

Over 300 Terabytes of event data have been transported with GridFTP. Using multiple streams in GridFTP has increased the throughput rate a factor of 5 over previous FTP technology. Matchmaking at the Grid scheduler level was developed for Condor-G as part of PPDG. Tens of thousands of jobs have been run using SAMGrid job planning and management tools based on Condor-G matchmaking. The failure rate of the grid job management infrastructure of <1% is very good and the end-to-end application failure rate is tolerable compared to that achieved in a non-grid local cluster environment.